

## Weekly Report (2017.4.10-2017.4.16)

### Done

1) Review records of previous meeting, try to understand meanings of data and find that I have some troubles on understanding them. Discuss with Mr. Tian's team on these problems, and have a deep understand on data.

2) Finish the design with designer.

3) Finish the requirements document which contains what need to be done and resolved. We still have some problem on communication. The requirements document is as follows:

#### 1 使用人群

本系统面向的使用人群是分析人员和管理员（指的是域中的各种管理员，在 csv 文件中有体现）。不同用户导入的文件默认是互不可见的，除非授权查看，才可以查看对方的文件。

#### 2 系统需求

##### 2.1 注册

记录使用者的用户名和密码

##### 2.2 登录

验证用户输入的用户名和密码是否正确

##### 2.3 数据处理与导入

###### 2.3.1 日志数据

在导入系统前，将四种格式的日志文件预处理为相同格式的文件，再导入系统中。预处理操作根据事件 id 分为两种情况来处理。事件 id 为 4624 和 4648 的事件为登录事件，需提取出用户名、IP、登录时间、登出时间，用于构建用户与计算机之间的关系图。非此两种 id 的其它事件需提取出 source address 与 destination address，用于统计出哪些 IP 段机器访问过此机器，就是通过分析，知道哪些段的机器是可以到这台机器的。

（**存在困难：**目前对方提供了一个不完备的提取工具 AnalyseEvent.exe，只能提取出登录事件的用户名和网关地址，对于非登录事件的信息提取功能和登录事件的登录、登出时间提取功能没有实现。未达到提取目标，有两种解决方案，一种是在对方提供的工具基础上进行修改，但是 C 语言编程不熟悉，另一种方案是，重新写脚本，但是我们手中的日志文件有限，也就造成了测试用例有限，不能保证适用于所有的日志文件提取。在尝试处理数据的过程中，由于是非专业人员，有很多地方不是很确认是不是应该这样做的，向对方询问，不能及时得到答复，影响开发进度，有时一次发很多问题，对方会漏掉几个问题不回答）

###### 2.3.2 csv 数据（域相关数据）

直接将 csv 文件导入到系统中，系统将文件中数据逐条读入内存，经过一定处理转存到 MySQL 数据库中。（对方说这里对效率没有要求，慢一点也没有关系）这里涉及到一个信任域的概念，例如 A 域与 B 域，域可以理解为公司，A 域是 B 域的信任域，而 B 域不是 A 域的信任域，那么，A 域管理员可以查看 B 域的文件，而 B 域管理员不能查看 A 域的文件。如果 A 域与 B 域互为信任域，则可互相查看文件，这里的查看文件，反映到系统中，就是可以搜索到对方的数据然后渲染出来，即查看数据渲染成图表以后的结果。

（**存在困难：**权限管理，计算机安全相关知识匮乏，数据库设计可能会存在漏

洞，有信息泄露的风险，例如 SQL 注入等）

### 2.3.3 其它 txt 文件

包括 DNS 数据、HOST 信息、端口扫描数据、端口信息，都需要通过处理存入数据库。由于企业内部开启了 DHCP 服务器，IP 会动态变化，所以不同时间下的 IP 会不同，DNS 文件中会有时间信息，端口扫描文件的时间信息为文件创建的时间。

## 2.4 用户管理与文件管理

密码修改

用户列表以及对应用户导入的文件列表

管理员预定义颜色含义

## 2.5 视图面板与信息面板

视图面板均提供缩略图与放大图，缩略图可放大查看。

### 2.5.1 计算机视图

将 csv 文件中 objectClass 为 computer 的信息绘制成树图，体现计算机的层次结构，由于节点数太多，采用合并叶子节点的方式进行绘制，即将拥有相同父节点的叶子节点合并到一起，用一个标有数字的大圆进行标记。提供的交互操作有：节点的单选、多选、自定义信息，其中单选和多选操作选择的节点的信息会展示在信息面板中。

### 2.5.2 用户视图

将 csv 文件中 objectClass 为 user 的信息绘制成树图，交互同计算机视图。

### 2.5.3 用户与计算机关系视图

利用处理过的日志文件、DNS 文件、端口探测文件绘制用户与计算机之间关系视图。由于节点数过多，在绘制前先进行聚类，此处有两种方案，一种是用户节点通过组数据进行聚类，计算机节点通过网关数据进行聚类（但是，网关数据是不完整的，对方已经肯定这种聚类方法可能对分析有意义）；另一种是抛开数据的内部意义，仅将他们看成节点，在二分图中寻找连通子图，进行聚类，此方案行得通，也许就能达到分析的目的。（我认为可能可以通过第二种方案的聚类，去完善网关信息，因为用户活动区域可能会有什么特点）提供的交互操作：点击连线，查看聚类内部的节点连接情况；分析人员根据经验分析出新的连线，可以手动添加连线。

### 2.5.4 计算机组视图

将 csv 文件中 objectClass 为 group 的信息与 objectClass 为 computer 的 memberOf 字段信息绘制出来，此处分两种视图，分两种角度进行展示，分别为使用 DN 字段绘制的层次图（树图），与使用 memberOf 字段绘制的包含关系图（tree map）。提供的交互操作：层次图交互同计算机视图，包含图的交互为通过点击查看更进一步信息。

### 2.5.5 用户组视图

将 csv 文件中 objectClass 为 group 的信息与 objectClass 为 user 的 memberOf 字段信息绘制出来，其余处理方式同计算机组视图

4) Discuss with Dr. Guo and Dr. Mei on the issue of rendering user-computer-graph. After the discussion, we propose a solution that cluster by connecting subgraphs before rendering the graph.

**To Do**

- 1) Process the data, implement the clustering function and rendering the user-computer-graph.
- 2) Keep in touch with Mr. Tian's team.